

修士論文の和文要旨

大学院電気通信学研究科 博士前期課程 情報通信工学 専攻 学籍番号 0330048

氏 名 柳澤 陽

論文題目 記号分解法を用いたデータ圧縮法の研究

要 旨

確率モデルに基づく無歪みデータ圧縮の標準的な1つのアプローチは、情報源から発生するシンボルの予測確率を逐次推定して、その予測確率を算術符号に適用するものである。無記憶情報源に対して、どんな未知の情報源に対しても平均冗長度を最小化するミニマックス最適符号としてよく知られている Krichevsky-Trofimov によって提案された確率予測法、Krichevsky-Trofimov (KT) 推定量は、文脈をベースとした確率予測法 (例えば、文脈木重み付け (CTW) 法) に用いられている。

しかし、テキストファイルのようなマルチアルファベット情報源に KT 推定量を適用した場合 (多進 KT 推定量) は、その性能は悪い。これを改善するために、シンボルを bit に分解して二進木に対応させ、各 bit の推定確率を求め、それらの積により次シンボルの予測確率を求める記号分解法アルゴリズムが Willems らによって提案されている。この方法と CTW 法を組み合わせたものは性能に改善が見られている。しかし、この記号分解法のアルゴリズムの導出及び性能の理論的な解析は示されていない。

本論文では、簡単化のために無記憶情報源を対象として、記号分解法アルゴリズムから仮定している確率モデル (パラメータ化) を解明し、そのモデルに対して仮定している事前分布を示す。次に無記憶情報源 (構造的アルファベット情報源) に対する記号分解法の冗長度解析を行う。さらに記号分解法のパラメータで多進 KT 推定量の事前分布である Jeffreys' prior を表し、記号分解法と多進 KT 推定量の prior の項の比較を行い、理解を深める。そしてシミュレーションとして、テキストファイルから統計をとり無記憶マルチアルファベット情報源を発生させ、記号分解法と多進 KT 推定量の冗長度を求め、記号分解法の優れた性能を示す。

最後に、記号分解法では各 bit の予測確率に 2 進 KT 推定量を用いているが、同じ bit の系列が発生する場合、その性能は悪くなってしまう。そこで同じ bit が続いてもその冗長度が数 bit で抑えられ、両方の bit が発生しても 2 進 KT 推定量と性能がほとんど変わらない Zero 冗長度推定量を各 bit の予測確率に用いる新しい方法を提案する。そしてこの方法の無記憶情報源に対する冗長度解析を行い、漸近的に最適な冗長度をもつことを示す。またシミュレーションにより、この方法と記号分解法との冗長度の比較を行う。